Audio Technical Metadata

- Importance of metadata to digital audio preservation
- Brief history of audio preservation standards
- Overview
  - Kinds of Audio Metadata
  - Categories of Audio Metadata
- AES57-2011 **AES standard for audio metadata—Audio object structures for preservation and restoration**
- **Closer look at Audio Technical Metadata**
- **Technical metadata / Structural Metadata**
- **Audio Object Manager: A real world example**

Metadata provides the framework for digital audio preservation.

Without metadata, digital audio preservation is not possible.

- Unless sound files possess a time stamp, there is no way to accurately determine the files' relationship to others in time.
- Without the structural metadata of an Audio Decision List (ADL), we cannot be certain of the transitions between multiple files in a single event.
- Without technical metadata about the files we cannot easily migrate them.
- Without information about how the digital audio was created and by whom, there is no way to judge what we have.

Standards for Digital Audio Preservation

THE LIBRARY OF CONGRESS
NATIONAL RECORDING PRESERVATION PLAN

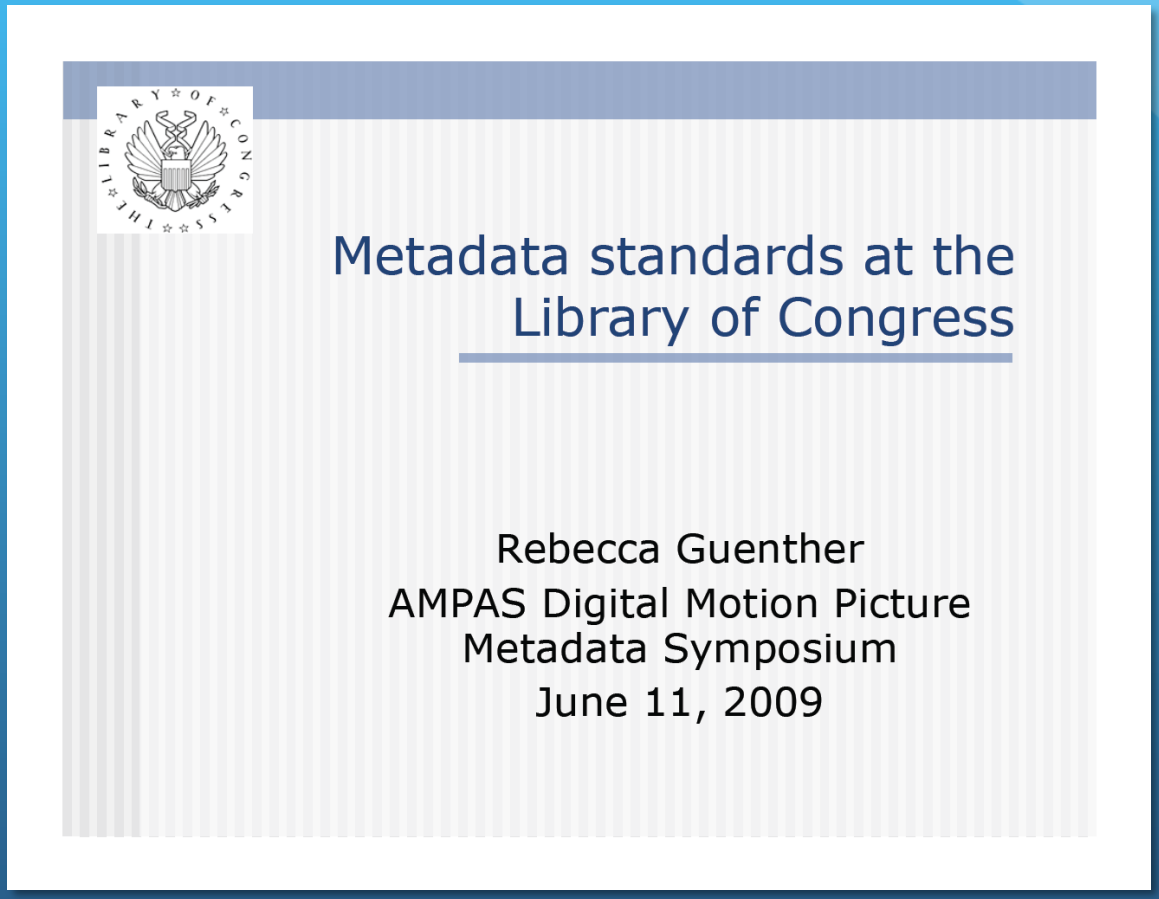
National Recording Preservation Act of 2000 [Public Law No: 106-474], SEC. 111. ESTABLISHMENT OF PROGRAM BY LIBRARIAN OF CONGRESS.

> *(a) IN GENERAL- The Librarian shall…implement a comprehensive national sound recording preservation program, in conjunction with other sound recording archivists, educators and historians, copyright owners, recording industry representatives, and others involved in activities related to sound recording preservation, and taking into account studies conducted by the Board.*

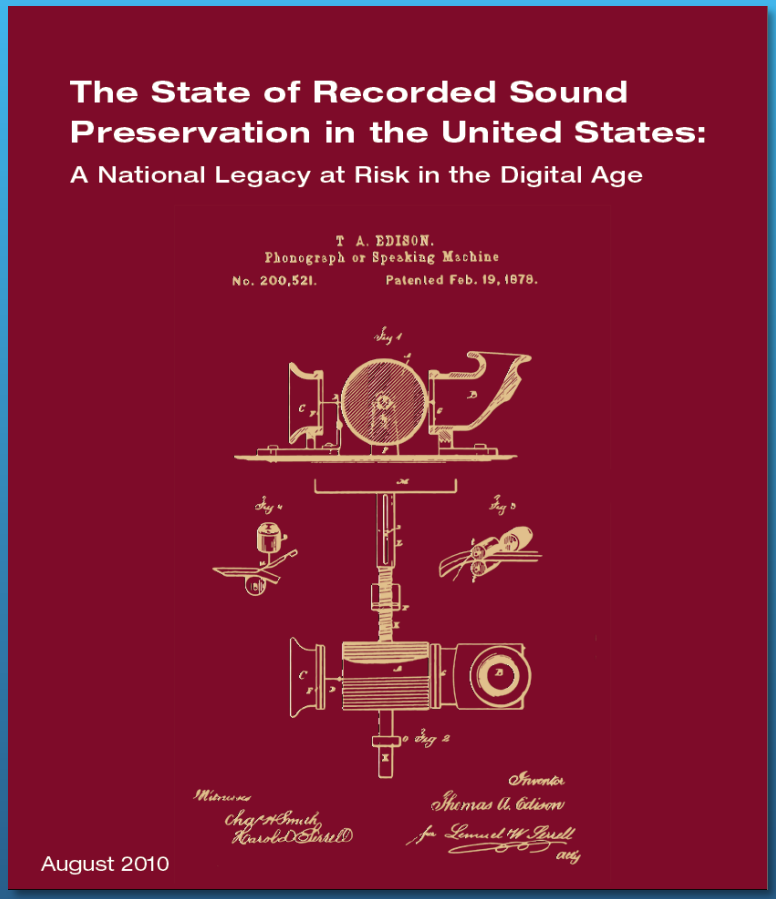
The National Recording Preservation Plan was prepared under the authority of the National Recording Preservation Act of 2000 [Public Law No: 106-474], which directed the Librarian of Congress "…to undertake studies and investigations of sound recording preservation activities…and recommend solutions to improve these practices."

Library of Congress National Recording Preservation Plan

Standards for Digital Audio Preservation

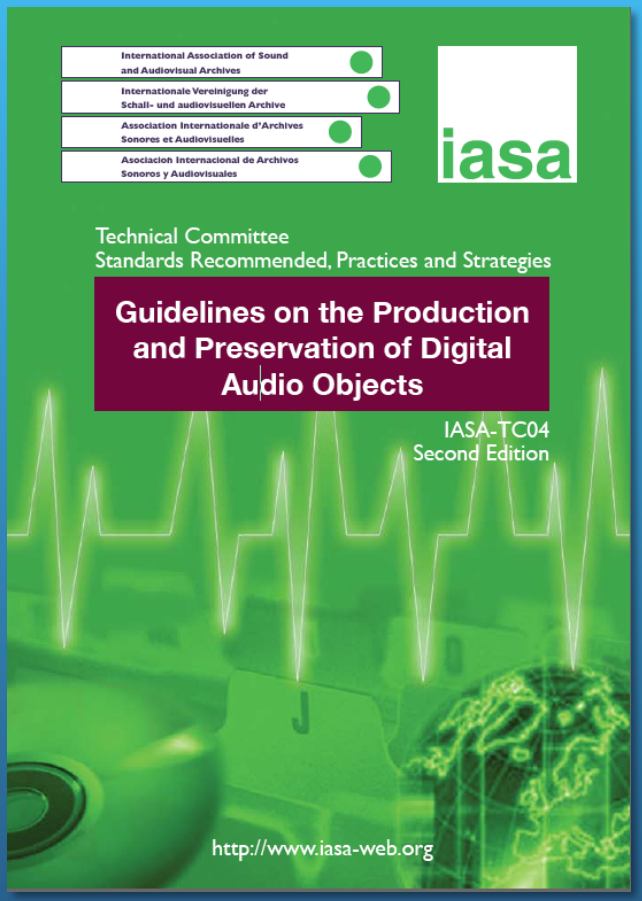


Standards for Digital Audio Preservation



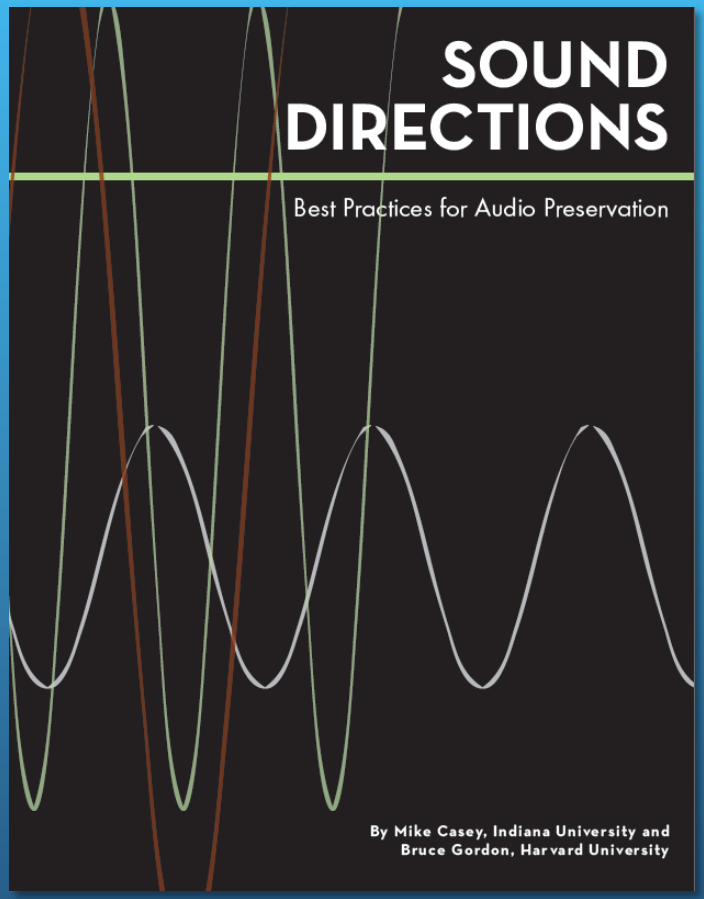
The State of Recorded Sound Preservation in the United States

Standards for Digital Audio Preservation



IASA TC-04
Guidelines on the Production and Preservation of Digital Audio Objects

Standards for Digital Audio Preservation



Sound Directions: Best Practices for Audio Preservation

Kinds of Audio Metadata

- Document the identity of a recording such as its title and call number,
- The performer's name,
- The occasion of the performance,
- The format of the original recording,
- How the original was played back.

Other forms of audio metadata

- Document the kinds of digital copies produced [i.e. Archival Master, de-noised Production Master, streaming Delivery Master],
- Where the digital files are stored,
- The relationships between the object's virtual and original manifestations and,
- How the digital files were made.

Categories of Audio Metadata

Descriptive Metadata	Cataloging data encoded in MARC or MODS
Administrative - Rights Management Metadata	Rights management metadata governs access to files. (The *Sound Directions* project did not address this topic.)
Administrative - Technical Metadata	Tape Speed, Oxide Coating, Groove Width, Sample Rate, Word Length, Coding, Noise Reduction, Condition Comments
Administrative - Digital Provenance Metadata	Process History (digiprov) for: archival transfer, sample rate conversion, de-noising or any DSP event, deliverable creation
Structural Metadata	AES31-3 Archival ADL, Audio Object Face, Region & Stream, BWF Time Stamp, PQ Marks, SMIL document, METS document (in its role documenting relationships using its <structMap>)

Descriptive Metadata

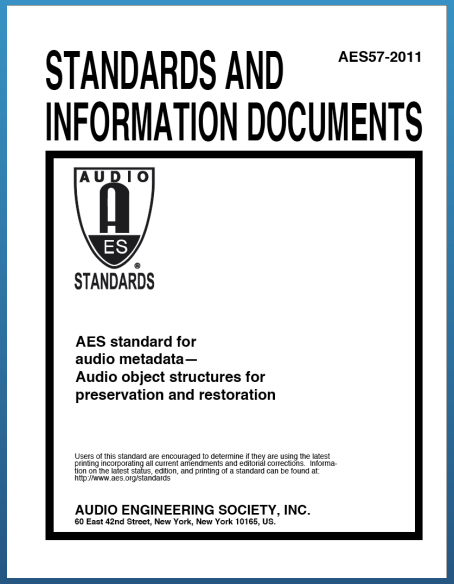
- Descriptive metadata is dedicated to curatorial information, rather than technical.
- This data identifies the object and its performances in a collections.

Rights Management Metadata

- Access rights and restrictions
- Preservation rights and restrictions

Technical Metadata

- Technical metadata describes the immediate technical attributes of a physical or file-based audio object including specifications that enable access to the content.
- AES57-2011



Digital Provenance Metadata

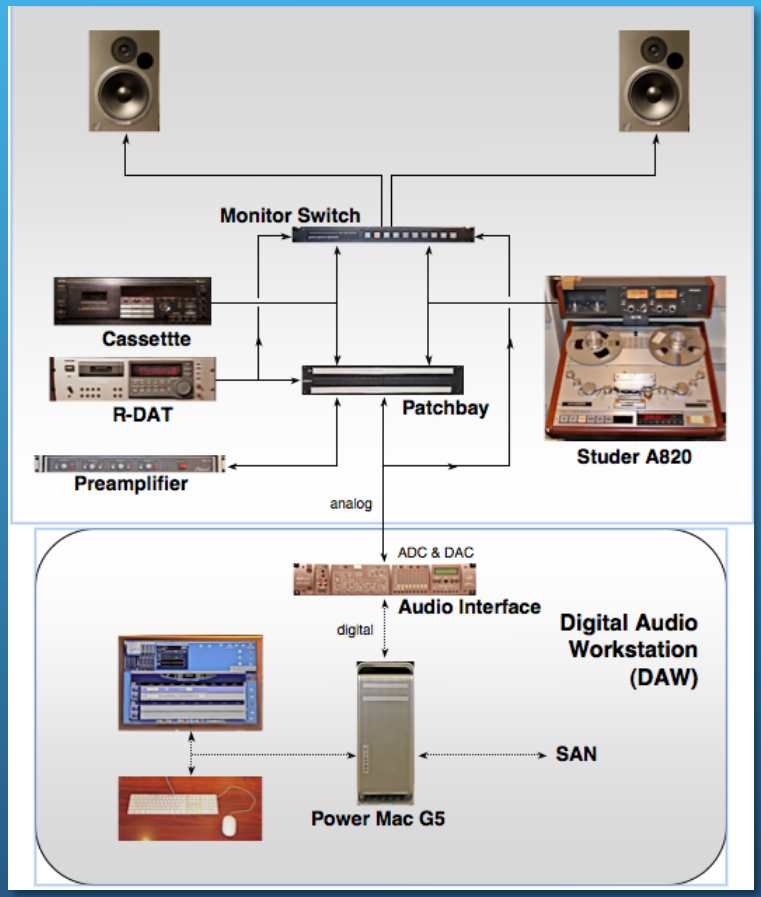
- Documents the “who, what, where, and how” of all preservation activity from transfer through the creation of deliverables.
- *A means to record preservation* processes for the distant future
- A vital tool in the more immediate preservation workflow—
  - facilitating teamwork and multiple project management by enabling recall of processes and settings at critical points in the workflow.
- It records the various components of each process such that any part can be evaluated and/or readily reproduced.

Digital Provenance Metadata

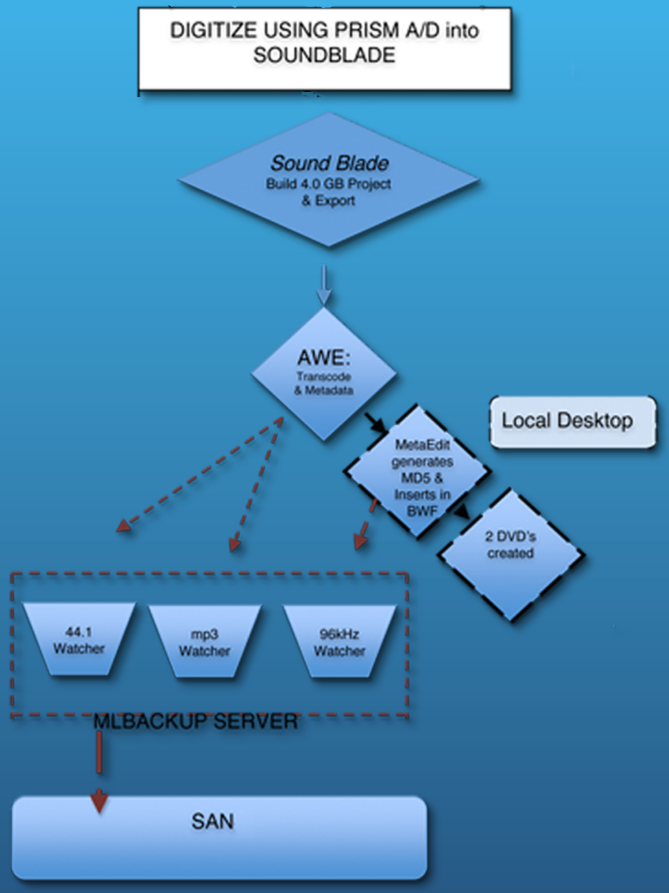
Further, each instance of processing of an audio object that has an input and an output is documented as an event.

- These events include such processes as
  - the archival transfer and AES31-3 ADL export,
  - any digital signal processing for production,
  - the sample rate and word length conversion during the creation of derivatives,
- Each event is listed with
  - the devices used,
  - all their inputs,
  - internal modules,
  - parameters and settings,
  - outputs,
  - along with the input media and the output media of the event.
- Auditing information attributing the process to
  - a specific operator,
  - location,
  - and time of each event.

Macaulay Library Analog Signal Path



Macaulay Library Archival sound file and derivatives



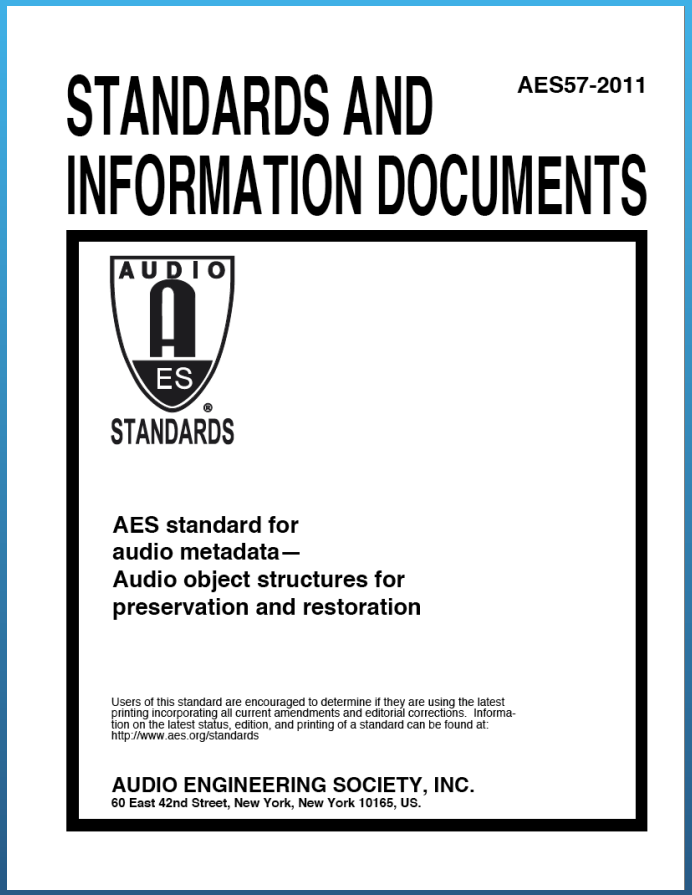
Structural Metadata

- It "ties the components of a complex or compound resource together and makes the whole usable."- Robin Wendler, "LDI Update: Metadata in the Library," *Harvard University Library Notes, no. 1286 (July/August* 1999), 4-5
- It can be used to express the relationships among objects.
- It can describe
  - the relative positions of the objects' streams on the carrier,
  - the boundaries of format based regions of physical objects,
  - and the mapping of the digital objects' audio streams through the use of pan automation data.
- It can also be used to define areas of interest within an audio object through the use of marker information.

Types of Structural Metadata

- Documentation of the source Audio Object structure
- AES31-3 Audio Decision List (ADL)
- Broadcast Wave Format time stamp
- Marker information defining areas of interest
- METS <structMap>

AES-57-2011



STANDARDS AND INFORMATION DOCUMENTS

AES57-2011

AUDIO ES STANDARDS

AES standard for audio metadata— Audio object structures for preservation and restoration

Users of this standard are encouraged to determine if they are using the latest printing incorporating all current amendments and editorial corrections. Information on the latest status, edition, and printing of a standard can be found at: http://www.aes.org/standards

AUDIO ENGINEERING SOCIETY, INC.
60 East 42nd Street, New York, New York 10165, US.

AES-57-2011

- This standard concerns the technical documentation of audio files and physical audio formats for long-term archival storage and preservation.
- This standard sets out the vocabulary to be used in describing digital and analog audio formats, including both those formats that exist in some tangible form such as a reel of tape and those that exist only as a set of bits, untied to a single audio carrier, such as a broadcast wave file.
- This vocabulary takes the form of an Extensible Markup Language (XML) schema. By doing so it provides a structured human readable instance document that can be easily parsed and manipulated using any of a number of freely available programming libraries and tools.

AES-57-2011

- The schema is designed for flexibility, providing a number of data elements that are optional.
- This schema provides for the collection of technical metadata in a number of broad categories, including:
  - Physical properties such as base material, oxide material, groove orientation, and track configuration;
  - Object dimensions such as height, width, depth, unwound length and shape;
  - Signal characteristics such as playback speed, sound field, and noise reduction;
  - Digital file characteristics such as audio data encoding, sample rate, bit depth, and byte order;
  - Condition—comments on preservation problems with the object.

Technical Metadata: Physical properties such as base material, oxide material, groove orientation, and track configuration;

- AES-57 : the **physicalProperties element is used to describe the physical characteristics of the audio** object.

- <format>audio cassette</format>
  - <physicalProperties>
    - <baseMaterial>Polyester</baseMaterial>

Various types of sound recording sources are included:

- Wire
- Cylinder
- Analog Tape
- Digital Tape
- Optical Disc
- Analog Disc

Technical Metadata: Object dimensions such as height, width, depth, unwound length and shape;

AES-57 : The <**dimensions**> **element may be used to provide physical measurements of the audio object.**

- <format>audio cassette</format>
- <dimensions>
  - <gauge unit=”inches”>0.125</gauge>
  - <length unit=»inches»>5304.4</length>
  - <thickness unit=»microns»>10.0</thickness>

Technical Metadata: Signal characteristics such as playback speed, sound field, and noise reduction;

AES-57 : The <speed> element may be used to indicate the playback speed of the described audio object or region.

<format>audio cassette</format>

<speed>

- <speedCoarse unit=»Inches per second»>1.875</speedCoarse>
- <varispeedAdjustment unit=»Percent»>0</varispeedAdjustment> </speed>

AES-57 : <trackLayout> element shall be used to describe the canonical track arrangement for tape-based audio objects.

- <track Layout>QUARTER-TRACK</trackLayout>

AES-57 : The <soundField> element shall be used to declare the overall soundstage to which the described audio object or region belongs. Valid values are MONO, STEREO, and SURROUND..

- <soundField>stereo</soundField>

Technical Metadata: Digital file characteristics such as audio data encoding, sample rate, bit depth, and byte order

AES-57 :
<audioDataEncoding> element shall be used to indicate the audio encoding scheme used when encoding occurred;

<sampleRate> element shall be used to indicate the sample rate of the audio data

<bitDepth> lement shall declare the number of bits per sample

<byteOrder> element shall be used to indicate the ‘endianess’ of the digital audio;

<audioDataEncoding>PCM</audioDataEncoding>
<sampleRate>S96000</sampleRate>
<bitDepth> 24</bitDepth>
<byteOrder> LITTLE_ENDIAN</byteOrder>

QUESTIONS?

Bill McQuay
wjm93@cornell.edu
607-254-1137

- AES57-2011
  - http://www.aes.org/
- IASA TC-04 Guidelines on the Production and Preservation of Digital Audio Objects
  - http://www.iasa-web.org/
- Sound Directions: Best Practices for Audio Preservation
  - http://www.dlib.indiana.edu/projects/sounddirections/

Structural Metadata: Bringing it all together
Source Audio Object Structure
Document Hierarchy

- Audio Object
- Face
- Region
- Stream

  The Face, Region and Stream sections can each be repeated in accordance with proper xml schema syntax.

Structural Metadata: Source Audio Object Structure

- Audio Object

  A general term used to describe digital files and physical audio carriers.

  It is the root of the hierarchy and contains metadata that pertains to the entire object.

  Each audio object is described by a single instance document in a strict one-to-one mapping.

Structural Metadata: Source Audio Object Structure

- Face

  A Face is a solitary stream, a sequence of solitary streams, or a group or sequence of groups of synchronous streams.

For example, a disc recorded on two sides contains two Faces —each side is a Face.

Structural Metadata: Source Audio Object Structure

- Region

  A Region is the partitioning of a Face based upon format.

  For example, if the first part of a Face is recorded at 15 ips, and the second part of the Face is recorded at 7.5 ips, then the Face would contain two Regions—one for each tape speed. A Face must have at least one Region.

Structural Metadata: Source Audio Object Structure

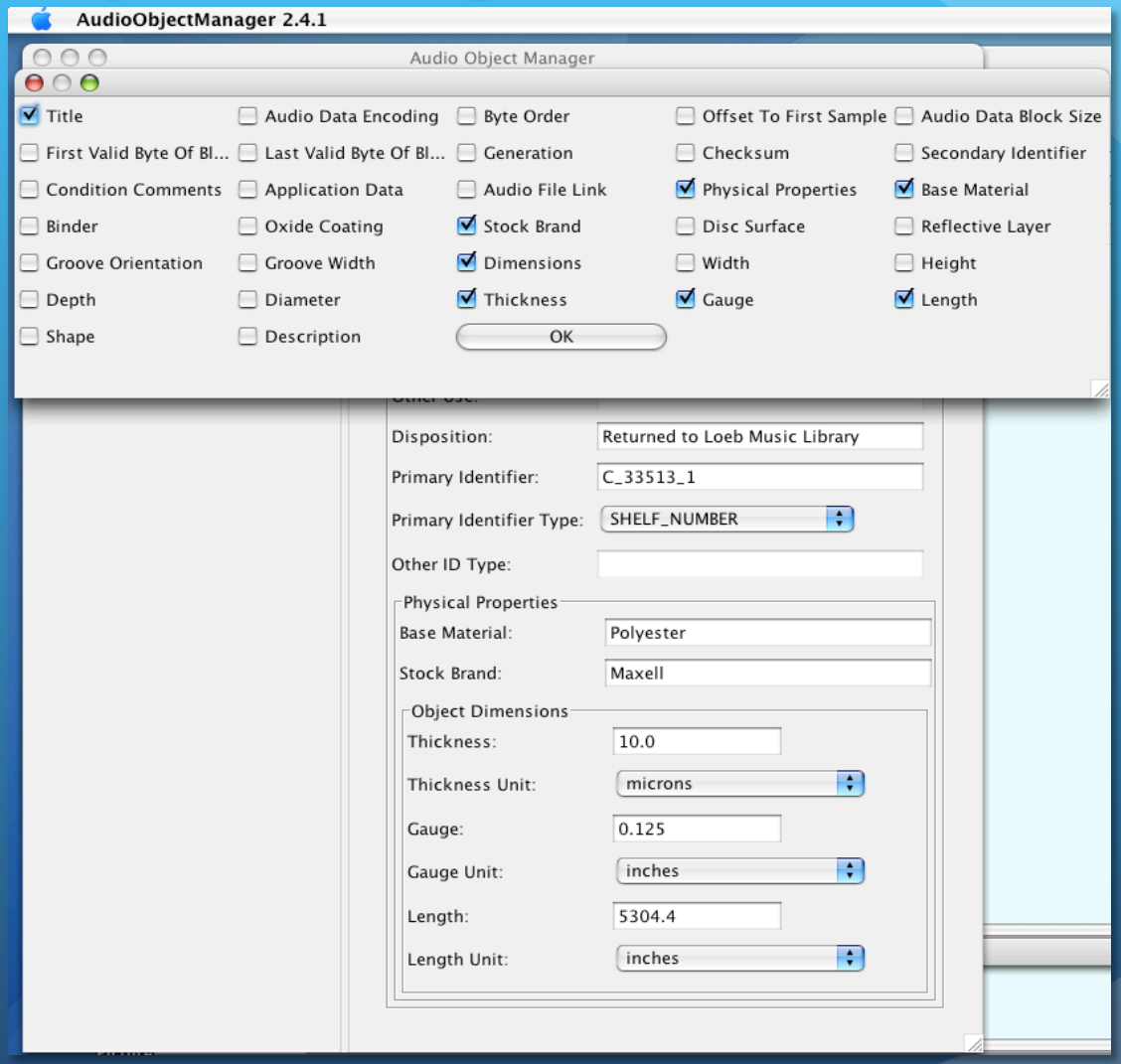
- Stream

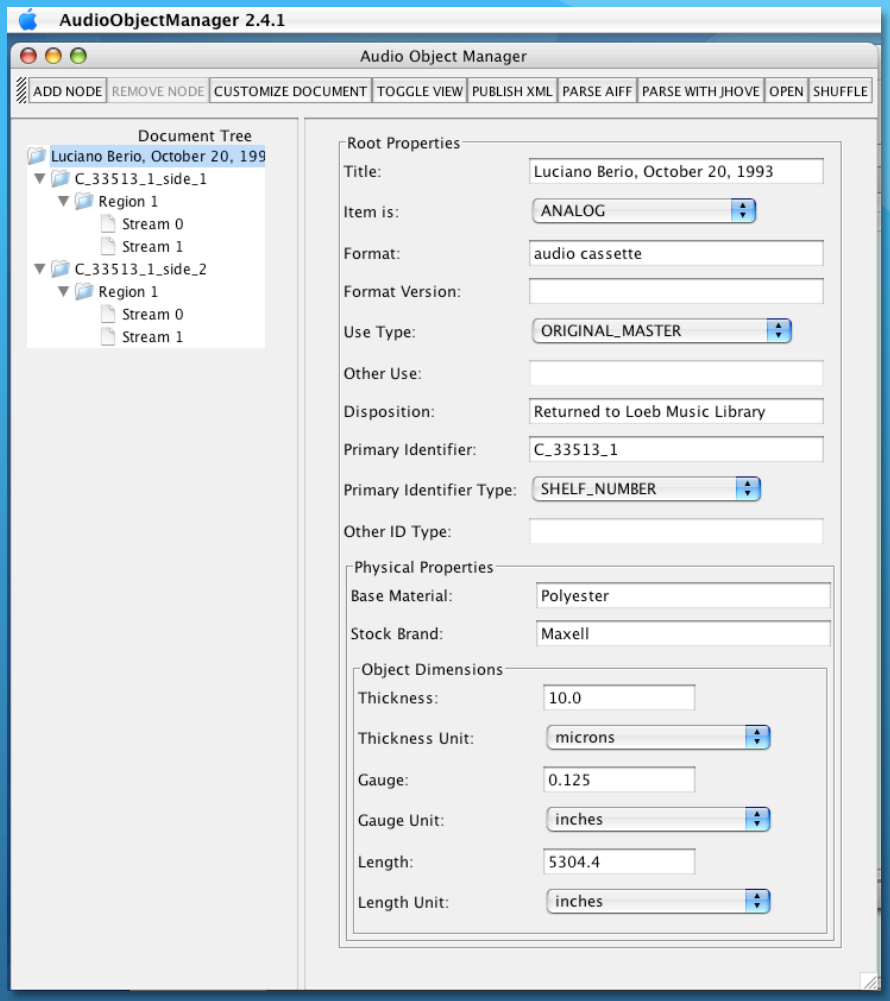
  A Stream is an individual channel of audio information contained within a Region. Each Region must have one or more Streams.

  Interleaved audio channels in a digital file are documented as separate individual Streams.

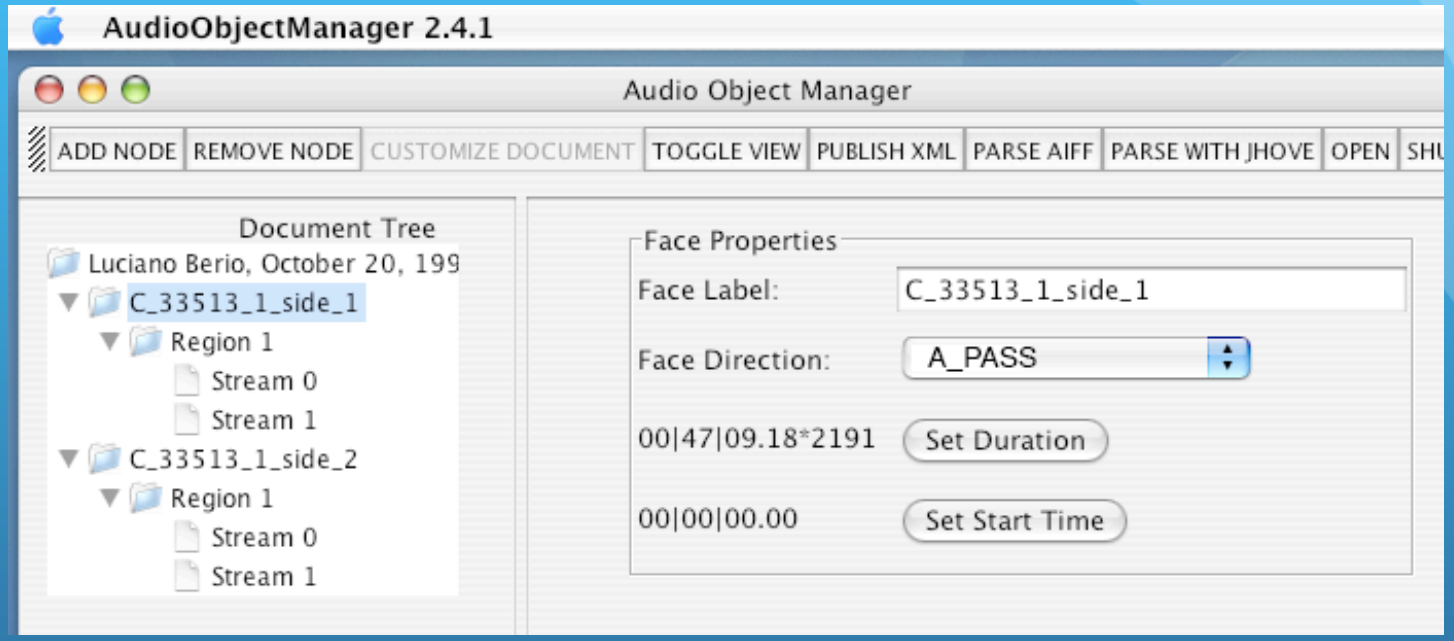
  For multiple audio files that represent multichannel audio, each file is a separate audio object.

Initial Configuration of AudioObjectManagerRoot

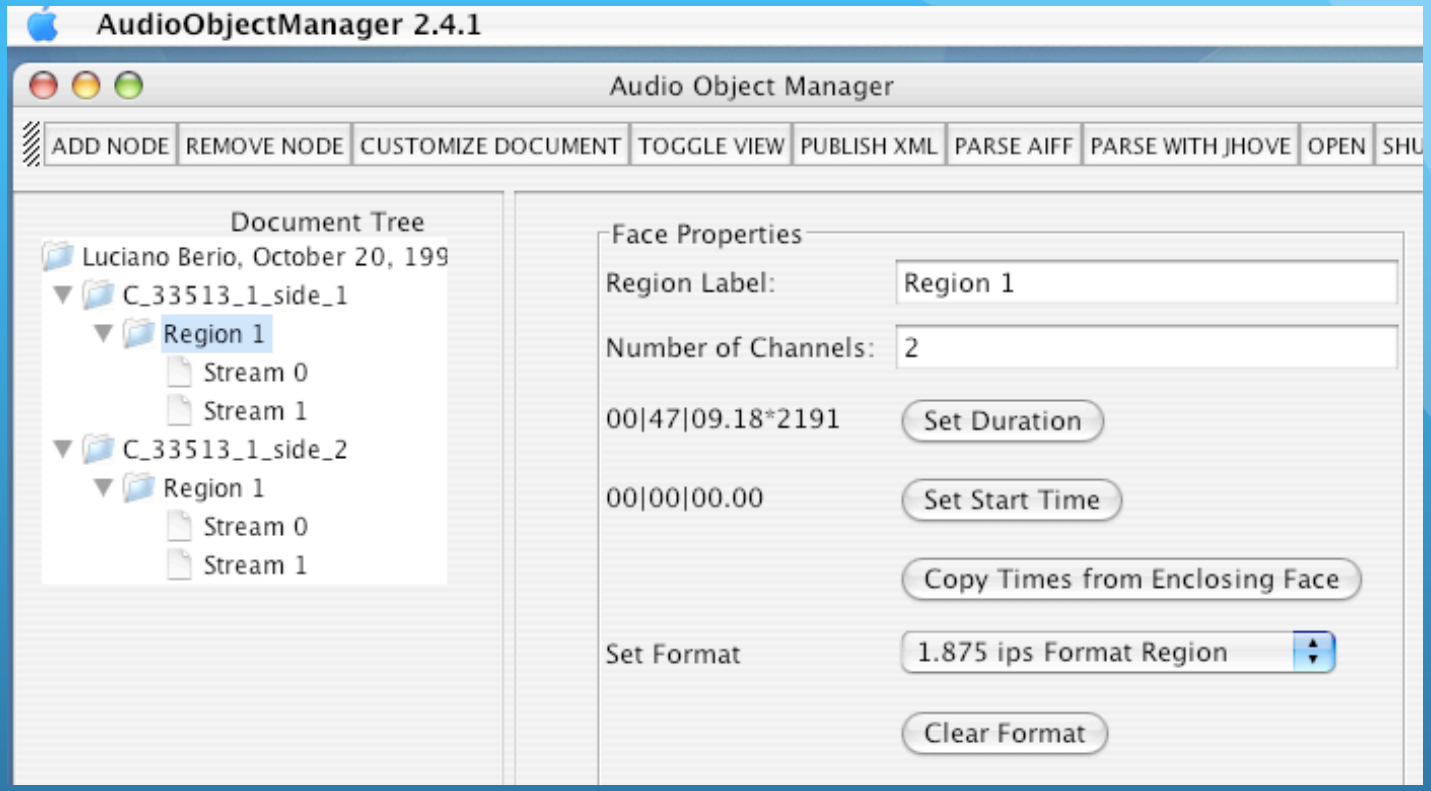




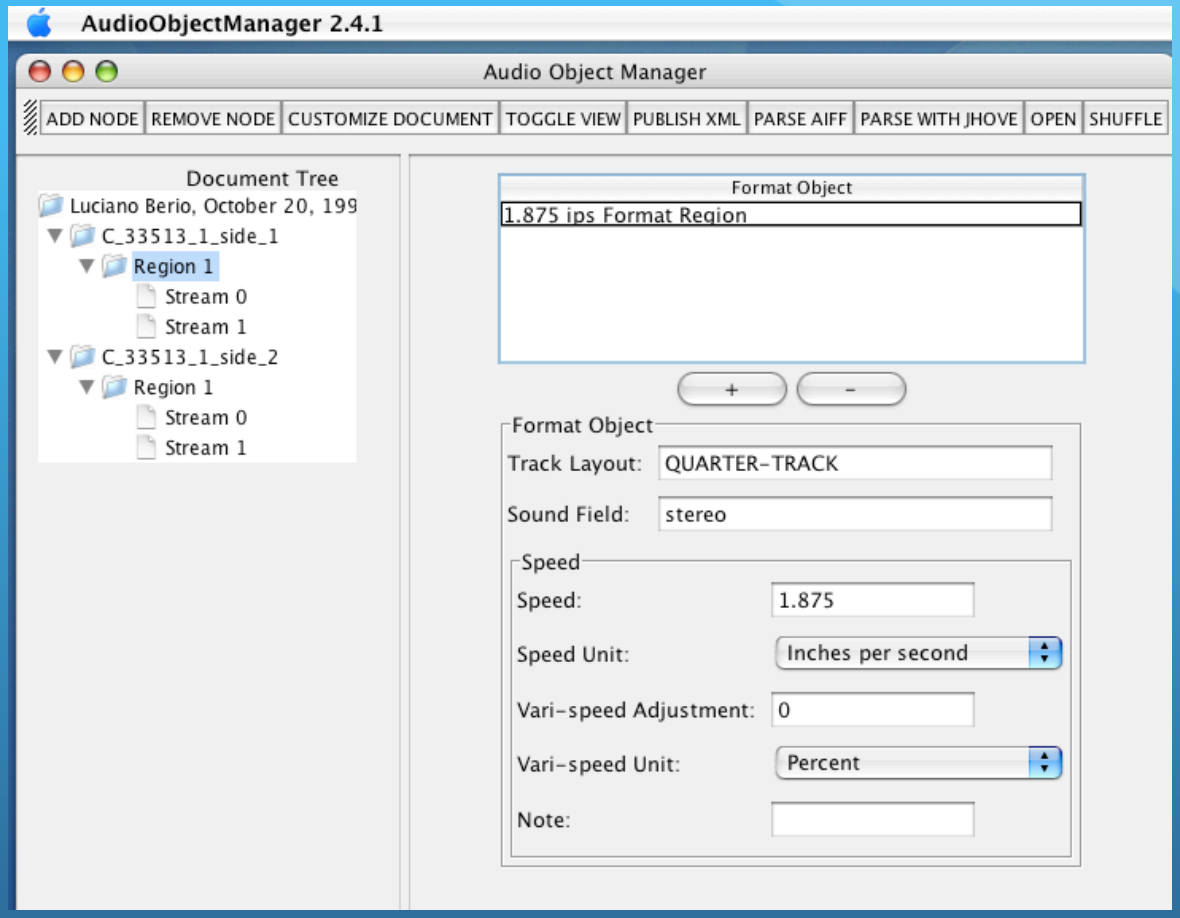
Root Properties



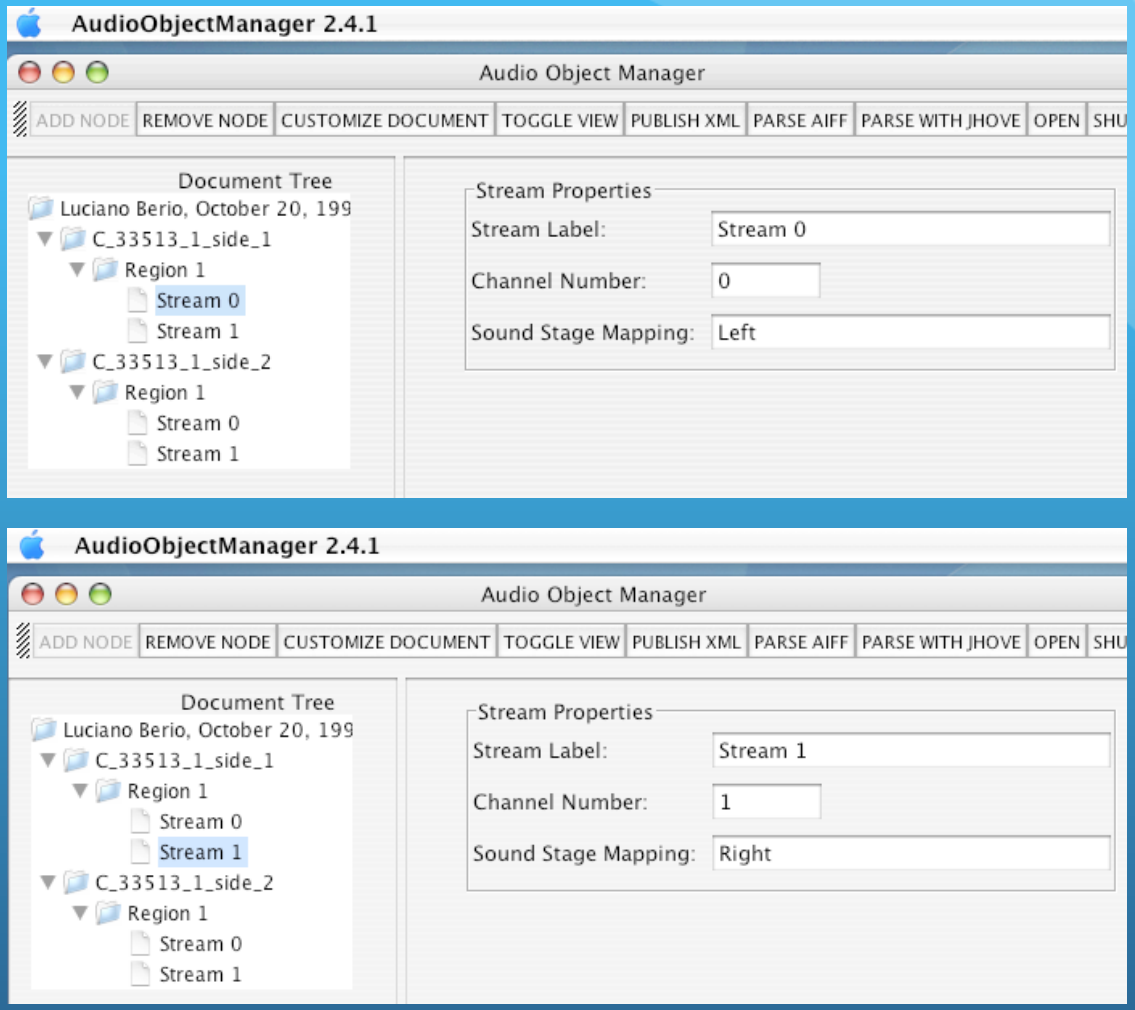
Face Properties



Region Properties



Configuring the Region's format



Stream Properties

Published document

```
<?xml version="1.0" encoding="UTF-8"?>
<audioObject xmlns:xlink="http://www.w3.org/1999/xlink"
   xmlns:tcf="http://www.aes.org/tcf"
   xmlns:xsd="http://www.w3.org/2001/XMLSchema"
   xmlns="http://www.aes.org/audioObject"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://www.aes.org/audioObject   http://hul.harvard.edu/ois/xml/xsd/drs/
audioObject.xsd"
   ID="_11559222219020.2609892689404265"
   title="Luciano Berio, October 20, 1993" analogDigitalFlag="ANALOG"
                disposition="Returned   to   Loeb   Music   Library"   schemaVersion="1.03b"
xsi:type="audioObjectType">
   <format>audio cassette</format>
   <physicalProperties>
      <baseMaterial>Polyester</baseMaterial>
      <stockBrand>Maxell</stockBrand>
      <dimensions>
         <gauge unit="inches">0.125</gauge>
         <length unit=»inches»>5304.4</length>
         <thickness unit=»microns»>10.0</thickness>
```

```
    </dimensions>
  </physicalProperties>
  <use useType=»ORIGINAL_MASTER»/>
  <primaryIdentifier identifierType=»SHELF_NUMBER»>C_33513_1</primaryIdentifier>
  <face ID=»_11559242914940.9381934972814039» direction=»A_WIND»
    audioObjectRef=»_11559222219020.2609892689404265» label=»C_33513_1_side_1»>
    <timeline>
      <tcf:startTime frameCount="30" timeBase="1000"
        videoField="FIELD_1" countingMode="NTSC_NON_DROP_FRAME">
        <tcf:hours>0</tcf:hours>
        <tcf:minutes>0</tcf:minutes>
        <tcf:seconds>0</tcf:seconds>
        <tcf:frames>0</tcf:frames>
        <tcf:filmFraming framing="NOT_APPLICABLE" xsi:type="tcf:ntscFilmFramingType"/>
      </tcf:startTime>
      <tcf:duration frameCount="30" timeBase="1000"
        videoField="FIELD_1" countingMode="NTSC_NON_DROP_FRAME">
        <tcf:hours>0</tcf:hours>
        <tcf:minutes>47</tcf:minutes>
        <tcf:seconds>9</tcf:seconds>
        <tcf:frames>18</tcf:frames>
        <tcf:samples sampleRate="S96000">
          <tcf:numberOfSamples>2191</tcf:numberOfSamples>
        </tcf:samples>
        <tcf:filmFraming framing="NOT_APPLICABLE" xsi:type="tcf:ntscFilmFramingType"/>
      </tcf:duration>
    </timeline>
    <region ID="_11559243209520.04244254843681039"
      formatRef="_11559243270550.5106679149708852"
      faceRef="_11559242914940.9381934972814039" label="Region 1">
      <timeRange>
        <tcf:startTime frameCount="30" timeBase="1000"
          videoField="FIELD_1" countingMode="NTSC_NON_DROP_FRAME">
          <tcf:hours>0</tcf:hours>
          <tcf:minutes>0</tcf:minutes>
          <tcf:seconds>0</tcf:seconds>
          <tcf:frames>0</tcf:frames>
          <tcf:filmFraming framing="NOT_APPLICABLE" xsi:type="tcf:ntscFilmFramingType"/>
        </tcf:startTime>
        <tcf:duration frameCount="30" timeBase="1000"
          videoField="FIELD_1" countingMode="NTSC_NON_DROP_FRAME">
          <tcf:hours>0</tcf:hours>
          <tcf:minutes>47</tcf:minutes>
          <tcf:seconds>9</tcf:seconds>
          <tcf:frames>18</tcf:frames>
          <tcf:samples sampleRate="S96000">
            <tcf:numberOfSamples>2191</tcf:numberOfSamples>
          </tcf:samples>
          <tcf:filmFraming framing="NOT_APPLICABLE" xsi:type="tcf:ntscFilmFramingType"/>
        </tcf:duration>
      </timeRange>
      <numChannels>2</numChannels>
      <stream ID=»_11559243815370.024875668740057644»
        label=»Stream 0» faceRegionRef=»_11559243209520.04244254843681039»>
        <channelAssignment channelNum=»0» mapLocation=»Left»/>
      </stream>
      <stream ID=»_11559243896970.36782358265958637»
        label=»Stream 1» faceRegionRef=»_11559243209520.04244254843681039»>
```

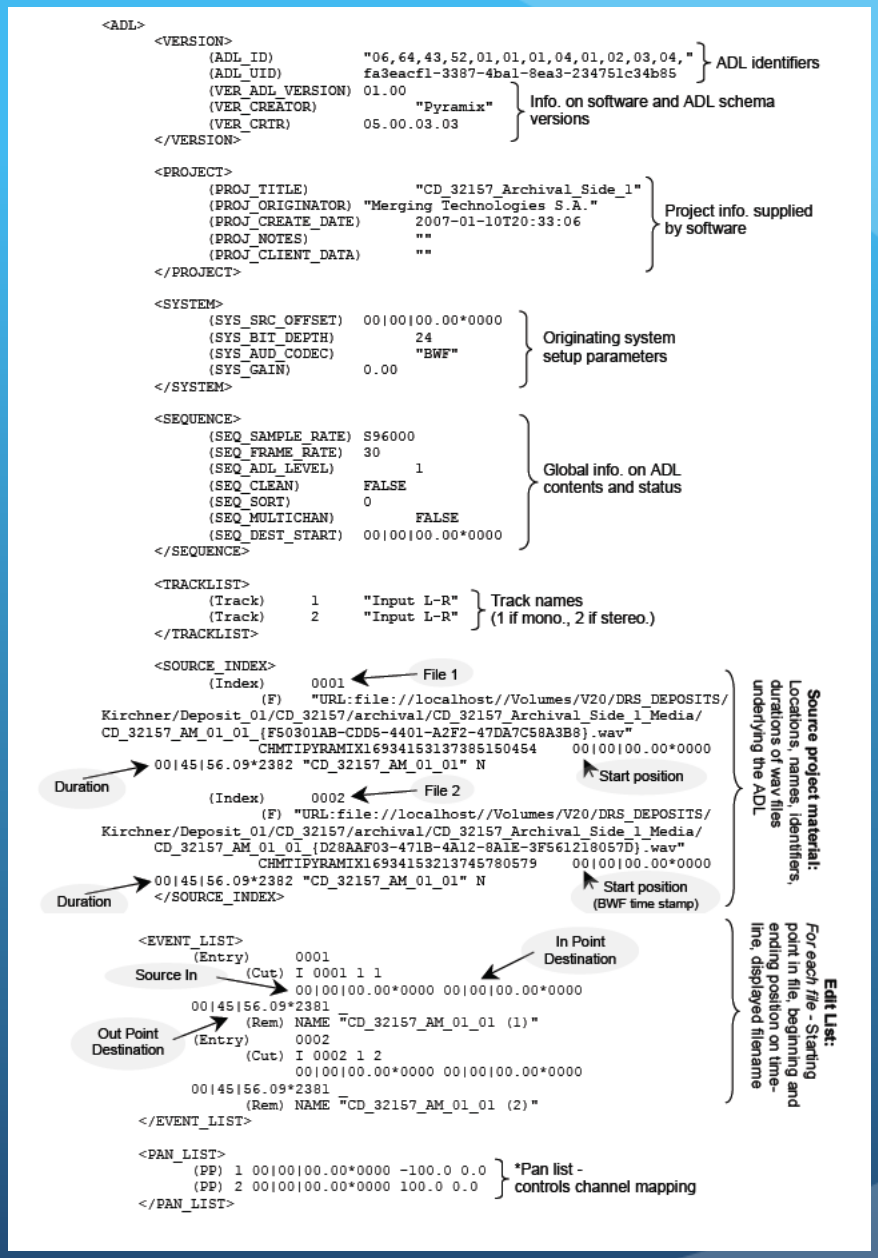
```
            <channelAssignment channelNum=»1» mapLocation=»Right»/>
          </stream>
        </region>
      </face>
      <face ID=»_11559245144540.03726796072569449» direction=»B_WIND»
        audioObjectRef=»_11559222219020.2609892689404265» label=»C_33513_1_side_2»>
        <timeline>
          <tcf:startTime frameCount=»30» timeBase=»1000»
            videoField=»FIELD_1» countingMode=»NTSC_NON_DROP_FRAME»>
            <tcf:hours>0</tcf:hours>
            <tcf:minutes>47</tcf:minutes>
            <tcf:seconds>12</tcf:seconds>
            <tcf:frames>18</tcf:frames>
            <tcf:samples sampleRate=»S96000»>
              <tcf:numberOfSamples>2191</tcf:numberOfSamples>
            </tcf:samples>
            <tcf:filmFraming framing=»NOT_APPLICABLE» xsi:type=»tcf:ntscFilmFramingType»/>
          </tcf:startTime>
          <tcf:duration frameCount=»30» timeBase=»1000»
            videoField=»FIELD_1» countingMode=»NTSC_NON_DROP_FRAME»>
            <tcf:hours>0</tcf:hours>
            <tcf:minutes>34</tcf:minutes>
            <tcf:seconds>18</tcf:seconds>
            <tcf:frames>4</tcf:frames>
            <tcf:samples sampleRate=»S96000»>
              <tcf:numberOfSamples>2339</tcf:numberOfSamples>
            </tcf:samples>
            <tcf:filmFraming framing=»NOT_APPLICABLE» xsi:type=»tcf:ntscFilmFramingType»/>
          </tcf:duration>
        </timeline>
        <region ID=»_11559245637120.8879793908058797»
          formatRef=»_11559243270550.5106679149708852»
          faceRef=»_11559245144540.03726796072569449» label=»Region 1»>
          <timeRange>
            <tcf:startTime frameCount=»30» timeBase=»1000»
              videoField=»FIELD_1» countingMode=»NTSC_NON_DROP_FRAME»>
              <tcf:hours>0</tcf:hours>
              <tcf:minutes>47</tcf:minutes>
              <tcf:seconds>12</tcf:seconds>
              <tcf:frames>18</tcf:frames>
              <tcf:samples sampleRate=»S96000»>
                <tcf:numberOfSamples>2191</tcf:numberOfSamples>
              </tcf:samples>
              <tcf:filmFraming framing=»NOT_APPLICABLE» xsi:type=»tcf:ntscFilmFramingType»/>
            </tcf:startTime>
            <tcf:duration frameCount=»30» timeBase=»1000»
              videoField=»FIELD_1» countingMode=»NTSC_NON_DROP_FRAME»>
              <tcf:hours>0</tcf:hours>
              <tcf:minutes>34</tcf:minutes>
              <tcf:seconds>18</tcf:seconds>
              <tcf:frames>4</tcf:frames>
              <tcf:samples sampleRate=»S96000»>
                <tcf:numberOfSamples>2339</tcf:numberOfSamples>
              </tcf:samples>
              <tcf:filmFraming framing=»NOT_APPLICABLE» xsi:type=»tcf:ntscFilmFramingType»/>
            </tcf:duration>
          </timeRange>
          <numChannels>2</numChannels>
```

```
        <stream ID=»_11559245736880.6900223089837544»
          label=»Stream 0» faceRegionRef=»_11559245637120.8879793908058797»>
          <channelAssignment channelNum=»0» mapLocation=»Left»/>
        </stream>
        <stream ID=»_11559245766480.932796829852367»
          label=»Stream 1» faceRegionRef=»_11559245637120.8879793908058797»>
          <channelAssignment channelNum=»1» mapLocation=»Right»/>
        </stream>
      </region>
    </face>
    <formatList>
      <formatRegion ID=»_11559243270550.5106679149708852»
        label=»1.875 ips Format Region» ownerRef=»_11559243209520.04244254843681039 _1
1559245637120.8879793908058797»>
        <speed>
          <speedCoarse unit=»Inches per second»>1.875</speedCoarse>
          <varispeedAdjustment unit=»Percent»>0</varispeedAdjustment>
        </speed>
        <trackLayout>QUARTER-TRACK</trackLayout>
        <soundField>stereo</soundField>
      </formatRegion>
    </formatList>
</audioObject>
```

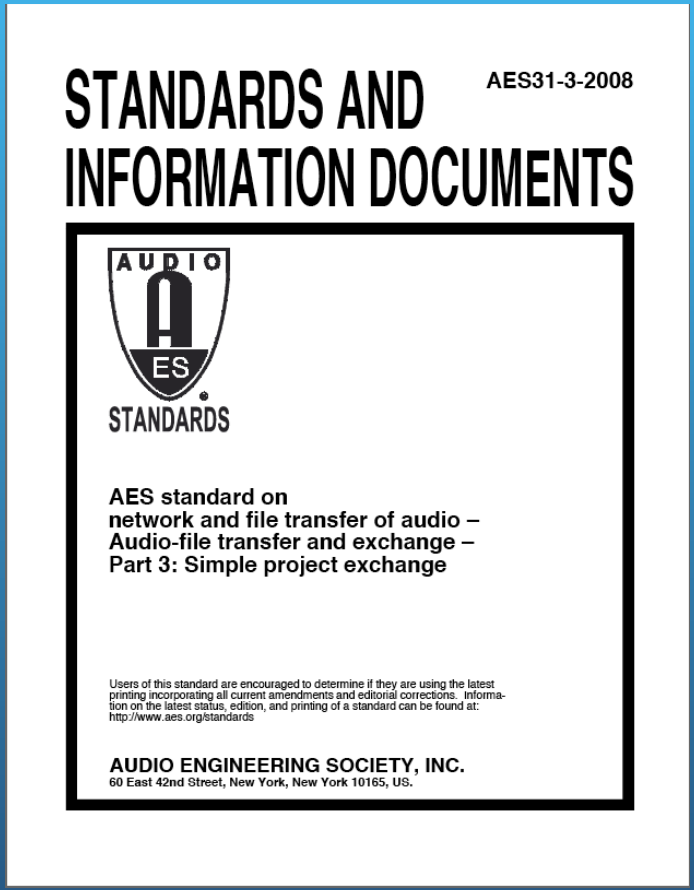
Bill McQuay
wjm93@cornell.edu
607-254-1137

- AES57-2011
  - http://www.aes.org/
- IASA TC-04 Guidelines on the Production and Preservation of Digital Audio Objects
  - http://www.iasa-web.org/
- Sound Directions: Best Practices for Audio Preservation
  - http://www.dlib.indiana.edu/projects/sounddirections/

An example of an AES31 ADL for side one of stereo audiocassette



Structural Metadata: Timelines



STANDARDS AND INFORMATION DOCUMENTS

AES31-3-2008

AUDIO ES STANDARDS

AES standard on
network and file transfer of audio –
Audio-file transfer and exchange –
Part 3: Simple project exchange

Users of this standard are encouraged to determine if they are using the latest printing incorporating all current amendments and editorial corrections. Information on the latest status, edition, and printing of a standard can be found at: http://www.aes.org/standards

AUDIO ENGINEERING SOCIETY, INC.
60 East 42nd Street, New York, New York 10165, US.

Structural Metadata: Timelines: ADL-Audio Decision List

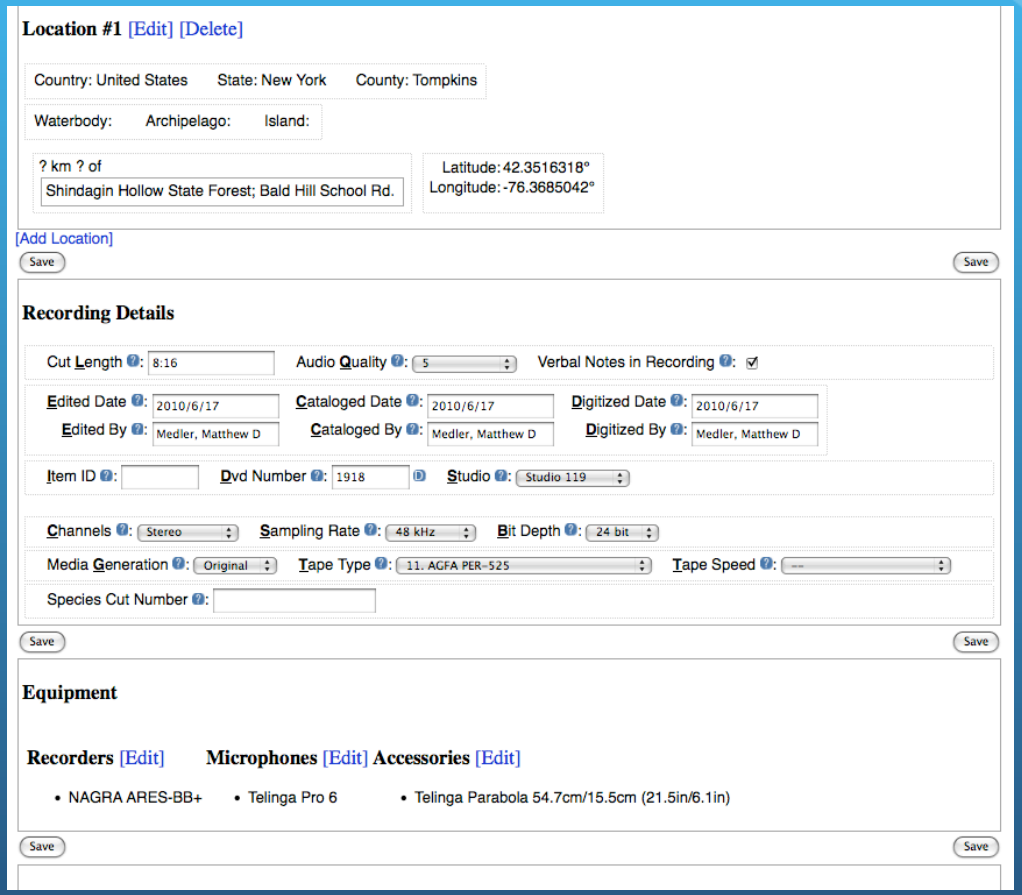
- An ADL may document the relationship between one Face of the source audio object and the digital files created from the preservation transfer of the Face.
- Using a pan list, an ADL can document the pan automation data for the streams of a Face. This data can be used by a DAW's software to position individual streams within the sound field of a project.
- An ADL can also document markers in reference to content in a digital file. These markers are used to define areas of interest in the content, and are often used to delineate performances

STANDARDS AND INFORMATION DOCUMENTS

AES57-2011

AES standard for audio metadata— Audio object structures for preservation and restoration

AUDIO ENGINEERING SOCIETY, INC.
60 East 42nd Street, New York, New York 10165, US.



Location #1 [Edit] [Delete]

Country: United States State: New York County: Tompkins

Waterbody: Archipelago: Island:

? km ? of
Shindagin Hollow State Forest; Bald Hill School Rd.

Latitude: 42.3516318°
Longitude: -76.3685042°

[Add Location]

Save Save

Recording Details

Cut Length: 8:16 Audio Quality: 5 Verbal Notes in Recording: ☑

Edited Date: 2010/6/17 Cataloged Date: 2010/6/17 Digitized Date: 2010/6/17
Edited By: Medler, Matthew D Cataloged By: Medler, Matthew D Digitized By: Medler, Matthew D

Item ID: Dvd Number: 1918 Studio: Studio 119

Channels: Stereo Sampling Rate: 48 kHz Bit Depth: 24 bit
Media Generation: Original Tape Type: 11. AGFA PER-525 Tape Speed: ---
Species Cut Number:

Save Save

Equipment

Recorders [Edit] **Microphones** [Edit] **Accessories** [Edit]

- NAGRA ARES-BB+
- Telinga Pro 6
- Telinga Parabola 54.7cm/15.5cm (21.5in/6.1in)

Save Save


BWF Information
Originator:
King, Ben F
Reference:
183
Date:
1977-02
Time:
Description:
macaulaylibrary.org/audio.do?id=183; Sabah, Malaysia; Sepilok; near Sandakan
Coding History:
Mono=1
OK

AES-57-2011

KIND	NAME	DATA TYPE	OCCURS		DEFAULT VALUE
			MIN	MAX	
ELEMENT	format	formatType	1	1	NONE
ELEMENT	physicalProperties	physicalPropertiesType	0	1	NONE
ELEMENT	appSpecificData	appSpecificDataType	0	Unbounded	NONE
ELEMENT	audioDataEncoding	string	0	1	NONE
ELEMENT	byteOrder	byteOrderType	0	1	NONE
ELEMENT	firstSampleOffset	long integer	0	1	NONE
ELEMENT	audioDataBlockSize	long integer	0	1	NONE
ELEMENT	firstValidByteOfBlock	long integer	0	1	NONE
ELEMENT	lastValidByteOfBlock	long integer	0	1	NONE
ELEMENT	use	useType	1	Unbounded	NONE
ELEMENT	primaryIdentifier	identifierType	1	1	NONE
ELEMENT	secondaryIdentifier	identifierType	0	Unbounded	NONE
ELEMENT	fileChecksum	checksumType	0	1	NONE
ELEMENT	soundDataChecksum	checksumType	0	1	NONE
ELEMENT	objectLink	locStringType	0	1	NONE
ELEMENT	face	faceType	1	Unbounded	NONE
ELEMENT	formatList	formatRegionListType	1	1	NONE
ELEMENT	objectCreationDate	dateTime	0	1	NONE
ATTRIBUTE	title	string	0	1	NONE
ATTRIBUTE	analogDigitalFlag	analogDigitalFlagType	1	1	NONE
ATTRIBUTE	generation	string	0	1	NONE
ATTRIBUTE	disposition	string	1	1	NONE
ATTRIBUTE	schemaVersion	string	1	1	1.0.0
ATTRIBUTE	*ID*	*ID*	*1*	*1*	*NONE*

The audioObject element may contain the following sub-elements and attributes.